Biophysical Journal Volume 106 April 2014 1489–1496

The Effects of Thermal Disorder on the Solution-Scattering Profiles of Macromolecules

Peter B. Moore*

Department of Chemistry, Yale University, New Haven, Connecticut

ABSTRACT The solution-scattering profiles of macromolecules are significantly affected by the thermal motions of their atoms, especially at wide scattering angles, even when only a single conformational state is significantly populated in solution. Here it is shown that the impact thermal motions have on the molecular component of the solution-scattering profile of a single-state macromolecule can be predicted accurately if the variances and covariances of the thermal excursions of its atoms from their average positions are known.

INTRODUCTION

Small angle x-ray solution scattering (SAXS) is an old, well-established technique for obtaining information about the conformation of biological macromolecules in solution whose time has finally come (see Glatter and Kratky (1)). For much of the field's history, the cost of data collection seemed, at best, to be higher than the value of information such data could provide. However, over the past two decades, the cost-benefit ratio of SAXS experiments has improved dramatically. The solution-scattering instruments now available at synchrotron light sources are vastly superior to the laboratory instruments used in the past. The data sets produced by synchrotrons and the newer point-focused laboratory instruments are largely free from systematic errors (such as slit smearing) that distorted the data collected in earlier years (see Graewert and Svergun (2)). In addition, data sets that took days to acquire using laboratory equipment can now be collected at synchrotrons in seconds. The cost to investigators has also fallen dramatically; access to the solution-scattering instruments at synchrotron facilities is provided free of charge. Not surprisingly, the number of articles published each year that include SAXS data has grown dramatically (2–4), and there has been a revival of interest in the underlying theory, e.g., Rambo and Tainer (5).

Solution-scattering experiments are often done to determine whether the conformations of biological macromolecules in solution are the same as the conformations they adopt in the crystalline state. These questions are answered by comparing experimental solution-scattering profiles with profiles that have been predicted for macromolecules using the coordinates of their atoms. In principle, profiles can be predicted using Debye's equation (6):

$$I(q) = \sum_{i=1}^{N}\sum_{j=i}^{N} f_i(q)f_j(q)\sin\left(qr_{ij}\right)/\left(qr_{ij}\right). \qquad (1)$$

Here $I(q)$ is proportional to the intensity of the scattered radiation observed at q, where q is $4\pi\sin\theta/\lambda$, θ is half the scattering angle, and λ is the wavelength of the radiation used. The value $f_i(q)$ is the scattering factor of the ith atom in the molecule, r_{ij} is the distance between the ith and jth atoms in the structure, and N is the number of atoms in the molecule.

It has long been known that even if the crystal structure of some macromolecule is exactly the same as its structure in solution, the profile predicted for it using Eq. 1 will not agree with experiment unless solvent effects are taken into account. Macromolecular solution-scattering profiles are differences between the scattering profiles of solutions of macromolecules and of the solvents in which they are dissolved. Macromolecular solutes occupy space in solutions that would otherwise be filled by solvent, and they alter the structure of the solvent in their immediate vicinities. The impact of these solvent perturbations on macromolecular solution-scattering patterns is large, especially at small scattering angles, but, fortunately, progress continues to be made in the development of algorithms for dealing with them (e.g., Svergun et al. (7), Park et al. (8), Schneidman-Duhovny et al. (9), Grishaev et al. (10), Poitevin et al. (11), and Liu et al. (12)).

It is less widely appreciated that even if solvent effects have been accurately accounted for, there are still likely to be measurable discrepancies between observed and predicted solution-scattering profiles, especially at wide angles. The reason is that even if the macromolecules in some solutions are all in the same conformational state, their structures will vary due to thermal fluctuations, and those variations will contribute to the solution-scattering profile observed (Makowski et al. (13)).

Many of the macromolecules studied by solution-scattering methods today adopt two or more distinctly different conformations in solution, or are flexible enough to display a more or less continuous range of conformations. A number of techniques have been developed both for predicting the effects that gross conformational heterogeneities have on solution-scattering profiles, and for extracting

Submitted August 9, 2013, and accepted for publication February 21, 2014.

*Correspondence: peter.moore@yale.edu

Editor: Lois Pollack.

0006-3495/14/04/1489/8 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.02.016

information about such heterogeneities from solution-scattering data (14–18). Some of them could be used to deal with the much smaller (~1 vs. ~10 Å) thermal variations in conformation characteristic of macromolecules in well-defined conformational states. For example: one might use molecular dynamics to generate an ensemble of all-atom models for the structure of some macromolecules at a specified temperature including the surrounding solvent, use Eq. 1 to estimate the solutions scattering profiles of all of the members of that ensemble, and then average those profiles to arrive at a final prediction (8).

Here we present a general description of the effect that small-amplitude thermal motions have on the nonsolvent component of solution-scattering profiles. It turns out that the type of information required to compute a specific component of the solution-scattering profile of a macromolecule corresponds to that needed to compute the diffraction pattern and the diffuse scattering patterns of crystals of the same macromolecule—namely, the coordinates of its atoms and the variances and covariances of their thermal motions (19).

THEORY AND RESULTS

The average scatter per molecule is not the same as the scatter of the average molecule

It is easy to show that the scattering profile of the average molecule in an ensemble differs from that of the ensemble to which it belongs. The following expression for the ensemble/time-averaged value of $I(q)$, $\langle I(q)\rangle$, emerges when Eq. 1 is expanded as a power series, and averages taken of all the structure-related variables it contains:

$$\langle I(q)\rangle = \sum_i\sum_j f_i(q)f_j(q)\left[1-\frac{q^2\left\langle r_{ij}^2\right\rangle}{3!}+\frac{q^4\left\langle r_{ij}^4\right\rangle}{5!}-\frac{q^6\left\langle r_{ij}^6\right\rangle}{7!}+\ldots\right]. \quad (2)$$

The difference in conformation between a particular molecule in an ensemble and its average member can be represented by adding to each vector between the ensemble/time-averaged position of the ith atom and the similarly averaged position of the jth atom, $\mathbf{R}_{ij}$, and a displacement vector, Δ_{ij}, that at any instant in time differs from one molecule to the next. (Note: $|\mathbf{R}_{ij}|$ $(= R_{ij})$ is not the same as the average distance between atoms i and j, $\langle|r_{ij}|\rangle$ (20).) To second-order in Δ_{ij}:

$$\left\langle|r_{ij}|\right\rangle = R_{ij}+\frac{\langle\mathbf{\Delta}_{ij}^2\rangle}{2R_{ij}}-\frac{\left\langle\left(\mathbf{R}_{ij}\cdot\mathbf{\Delta}_{ij}\right)^2\right\rangle}{8R_{ij}^3}.$$

For any given pair of atoms within any molecule in the ensemble, the length and direction of Δ_{ij} will vary with time. However, at any instant in time, the ensemble average value of all the Δ_{ij} values will be zero, and for any specific molecule, their time-averaged values will all be zero also. Thus for any member of the ensemble at any instant in time,

$$r_{ij}^2 = \mathbf{R}_{ij}^2+2\mathbf{R}_{ij}\cdot\mathbf{\Delta}_{ij}+\mathbf{\Delta}_{ij}^2,$$

and so on for higher-order terms. Furthermore, the ensemble/time average of $r^2{}_{ij}$, $\langle r^2{}_{ij}\rangle$, is

$$\left\langle r_{ij}^2\right\rangle = R_{ij}^2+\left\langle\Delta_{ij}^2\right\rangle,$$

and the ensemble/time-average value of $\langle r^4{}_{ij}\rangle$ is

$$\left\langle r_{ij}^4\right\rangle = R_{ij}^4+\left\langle\Delta_{ij}^4\right\rangle+2R_{ij}^2\left\langle\Delta_{ij}^2\right\rangle+4R_{ij}^2\left\langle\Delta_{ij}^2\right\rangle\langle\cos^2\theta\rangle,$$

where θ is the angle between $\mathbf{R}_{ij}$ and Δ_{ij}. It is obvious that the expression that emerges when these average values are substituted into Eq. 2 is not the same as

$$I(q) = \sum_i\sum_j f_i(q)f_j(q)\sin\left(qR_{ij}\right)/\left(qR_{ij}\right),$$

which is the solution-scattering profile of the average member of the ensemble.

The Debye equation can be modified to take thermal motions into account

At any instant, the intensity of the x-rays scattered by a particular member of an ensemble of chemically identical macromolecules in some direction relative to that molecule, $I(\mathbf{q})$, is proportional to

$$I(\mathbf{q}) = \sum_{i=1}^{N}f_i^2(q)+\sum_{i=1}^{N}\sum_{j\neq i}^{N}f_i(q)f_j(q)\times\exp\left(-i\mathbf{q}\cdot\left(\mathrm{R}_{ij}+\mathbf{\Delta}_{ij}\right)\right), \quad (3)$$

where $\mathbf{q}$ is a vector in reciprocal space of length q, parallel to the difference between a unit vector pointing in the direction the scattering is being observed and a unit vector pointing in the direction of the incident radiation (21). The solution-scattering profile of an ensemble of such molecules is the time/ensemble average of the profile obtained by rotationally averaging the single molecule scattering pattern described by Eq. 3 in one atom pair at a time:

$$I(q) = \sum_{i=1}^{N}f_i^2(q)+\sum_{i=1}^{N}\sum_{j\neq i}^{N}f_i(q)f_j(q)\times\left\langle\frac{1}{4\pi}\int_0^{\pi}\sin\theta d\theta\int_0^{2\pi}\exp\left(-iq\cdot\left(\mathbf{R}_{ij}+\mathbf{\Delta}_{ij}\right)\right)d\varphi\right\rangle_{\mathrm{ens}}. \quad (4)$$

In this equation, θ is the angle between $\mathbf{q}$ and $\mathbf{R}_{ij}$ (not one-half the scattering angle), and φ is the azimuthal angle of $\mathbf{R}_{ij}$

in any convenient, local coordinate system, one axis of which is aligned with $\mathbf{R}_{ij}$.

As the derivation provided in the Supporting Material shows, integrals that are the products of $\mathbf{R}_{ij}$ and Δ_{ij}, which emerge after the exponent term in Eq. 4 is written, are straightforward to evaluate. The expressions obtained can be further simplified if the distribution functions for the displacements of atoms from their average positions can all be assumed to be at least approximately Gaussian, and the three-dimensional distribution function for each pair of atoms, $p_{ij}(x,y,z)$, can be represented as the product of three independent distributions, i.e., if

$$p_{ij}(x,y,z) = p_{ijx}(x)p_{ijy}(y)p_{ijz}(z).$$

If these conditions are met, then

$$\mathbf{C}_{ij}^{T}\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}$$

is the component of the variance of the fluctuations in the positions of atoms i and j that is parallel to $\mathbf{R}_{ij}$. Because this component of their fluctuations will have a bigger impact on the distance between them than any other, it is not surprising that it determines the magnitude of the first-order correction that has to be applied to the Debye equation to account for thermal motions. The quantity

$$\left(Tr\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)-3\mathbf{C}_{ij}^{T}\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}\right)$$

is also easy to understand. Its value will differ from zero only if the variance of the mutual motions of a pair of atoms is anisotropic. It is thus a first-order correction for anisotropy. The second double-sum term is a second-order anisotropy correction. (Equation 4 can also be obtained by expanding Eq. 1 as a Taylor series, and then evaluating its terms out to fourth-order.)

$$\begin{aligned}\langle I(q)\rangle \approx & \sum_i f_i(q)^2 + 2\sum_i\sum_{j>i} f_i(q)f_j(q)\times\left[\frac{\sin\left(qR_{ij}\right)}{\left(qR_{ij}\right)}+(A/2)\left[Tr\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)-3\mathbf{C}_{ij}^{T}\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}\right]\right]\\ &\times\exp\left(-\frac{1}{2}q^2\mathbf{C}_{ij}^{T}\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}\right)-\sum_{i=1}^{N}\sum_{j>i}^{N} f_i(q)f_j(q)\left(\frac{q^2}{8R_{ij}^2}\right)\left(\frac{\sin\left(qR_{ij}\right)}{qR_{ij}}+\frac{3A}{q^2}\right)\\ &\times\left[Tr\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)-3\mathbf{C}_{ij}^{T}\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}\right]\left[3Tr\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)-5\left(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\right)\mathbf{C}_{ij}\right].\end{aligned} \tag{5}$$

Here $\mathbf{C}_{ij}$ is a three-dimensional column vector, the components of which are

$$\left(\cos\theta_{ijx},\ \cos\theta_{ijy},\ \cos\theta_{ijz}\right),$$

the direction cosines of $\mathbf{R}_{ij}$, with respect to the coordinate system, are used to specify atomic positions, and $\mathbf{R}_{ij}{}^{\mathrm{T}}$ is its transpose. A is defined as follows:

$$A\equiv\frac{\cos\left(qR_{ij}\right)}{R_{ij}^2}-\frac{\sin\left(qR_{ij}\right)}{qR_{ij}^3}.$$

$\mathbf{V}_i$ is a 3×3 matrix, the elements of which are

$$\begin{matrix}\langle\delta_{ix}\delta_{ix}\rangle & \langle\delta_{ix}\delta_{iy}\rangle & \langle\delta_{ix}\delta_{iz}\rangle\\ \langle\delta_{iy}\delta_{ix}\rangle & \langle\delta_{iy}\delta_{iy}\rangle & \langle\delta_{iy}\delta_{iz}\rangle\\ \langle\delta_{iz}\delta_{ix}\rangle & \langle\delta_{iz}\delta_{iy}\rangle & \langle\delta_{iz}\delta_{z}\rangle,\end{matrix}$$

which is the variance-covariance matrix for the displacements of the ith atom in the molecule expressed in the relevant coordinate system. $\mathbf{V}_j$ is the corresponding matrix for the jth atom, and $\mathbf{V}_{ij}$ represents the covariances of the displacements of atoms i and j. Its elements have the form $\langle\delta_{ix}\delta_{jx}\rangle$. Tr, first seen in Eq. 4, is the trace of whatever matrix it is applied to.

Equation 4 is less formidable than it looks, where

Equation 4 is usefully accurate

Equation 4 is an approximation to the first few terms of an infinite series. Are the predictions it produces accurate enough over a wide-enough range of scattering angles to be interesting? This question was addressed by generating a large number of solution-scattering data sets in silico, each of which approximated the solution-scattering profile that would be obtained from an ensemble of diatomic molecules of predetermined average bond length, the atoms of which move thermally. The distribution functions describing the excursions of atoms from their average locations were all products of three orthogonal Gaussians, the variances of which could be altered at will. A random number generator was used to produce 10,000 examples of molecules drawn from the relevant distributions for each diatomic molecule. The solution-scattering profiles of all the molecules in each ensemble were computed separately, and then averaged to yield the solution-scattering profile of the ensemble as a whole. (Tests done with data sets generated using smaller numbers of molecules

demonstrated that 10,000 profiles were more than enough to ensure convergence.) These simulated profiles were then compared with the profiles predicted for the corresponding molecules using both Eq. 4, which takes thermal disorder into account, and the Debye equation, which does not.

Not surprisingly, for fixed $\langle \Delta_{ij}^2 \rangle$, the impact that fluctuations have on average solution-scattering profiles is large when the average separation between atoms is small, and falls rapidly with increasing separation. It is easy to understand why this is so. The average solution-scattering profile for a pair of atoms separated by a distance of 20 ± 1 Å is much closer to that of a pair of atoms separated by exactly 20 Å, than is the average solution-scattering profile of a pair of atoms separated by a distance of 4 ± 1 Å to that of a pair of atoms separated by exactly 4 Å.

Fig. 1, top, shows some results obtained for a pair of atoms having an average separation of 3 Å, both of which are fluctuating in position isotropically such that $\langle \Delta_{ij}^2 \rangle = 1.71$ Å^2. (This value for $\langle \Delta_{ij}^2 \rangle$ was chosen so that the results shown in Fig. 1 would correspond to those displayed in Fig. 2 (see below).) The quantities plotted are the simulated profiles divided by either the profile predicted using the Debye equation (*dotted line*) or the profile predicted using Eq. 4 (*solid line*). The larger the deviation from 1.0, the poorer the match is between simulation and prediction. Fig. 1, bottom, is a similar comparison. In this case, however, while $\langle \Delta_{ij}^2 \rangle$ and the average distance between atoms was the same, only motions perpendicular to the vector between the two atoms were allowed. In both cases, as in all the other cases tested, the superiority of the profiles obtained using Eq. 4 is obvious. Equation 4 appears capable of producing usefully accurate predictions out to $q \sim 1.0$.

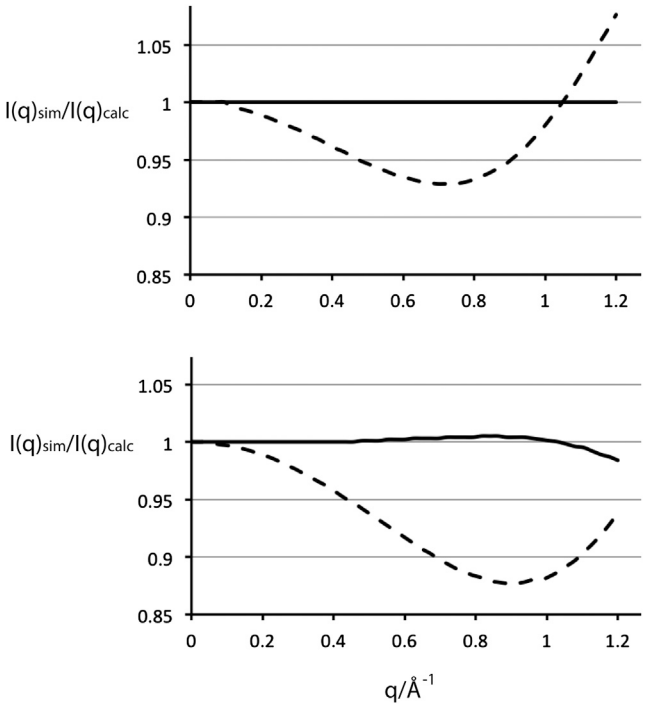

FIGURE 1 The effect of thermal disorder on two-atom solution-scattering profiles. The contribution made by a single atom to the scattering profile of a diatomic molecule having a specified average interatomic separation, but subject to thermal motions, was obtained by averaging the profiles computed for a large number of different molecules taken at random from the distribution of molecules possible for that diatomic, given the model specified for the thermal motions of its atoms. The distribution functions used for atomic motions were all triaxial Gaussians, but their variances in different directions could be varied. For each such computation, the number of scattering profiles averaged was 10,000. These simulated profiles, $1 + \langle \sin(qr_{ij})/qr_{ij} \rangle$, were compared to the corresponding profile predicted for the population using the Debye equation, $1 + \sin(qR_{ij})/qR_{ij}$, with R_{ij} being the ensemble average value of r_{ij}, and the corresponding profile obtained using Eq. 4. What is plotted in both panels is the profile obtained from dividing the simulated profile by the computed profile. (*Solid lines*) Ratio of the simulated profile to the Eq. 4 profile. (*Dashed lines*) Ratio of the simulated profile to the Debye equation profile. (*Top panel*) The average distance between atoms is 3.0 Å, and the thermal motions of both atoms are isotropic. The variance of the displacements for both atoms is 1.71 Å^2, which implies that B-factor for both is 45 Å^2 (see Fig. 2 and its discussion). (*Bottom panel*) The average distance between atoms is again 3.0 Å, but their thermal motions are highly anisotropic. No thermal motions are allowed along the vector joining the two atoms; all of the thermal motions permitted are in directions orthogonal to that bond. However, the variance of the displacements of both atoms remains 1.71 Å^2.

Crystallographic B-factors convey information about atomic displacements

Interatomic distances, which can easily be gleaned from the information about macromolecular structures available in databases like the Protein Data Bank (PDB), are all that is needed to predict solution-scattering profiles of macromolecules of known structure using the Debye equation. The extra information required to make predictions using Eq. 4 are the variance/covariance matrices that characterize the thermal motions of the atoms in a macromolecule. Although it is far from obvious where that information is to be found, at least an approximation to some of it is routinely reported for crystal structures.

Atomic displacements are relevant to crystallography because they directly affect the intensities of the Bragg reflections that must be measured to solve structures. The Fourier transform of the average electron density distribution in the unit cells of some crystal, $F(\mathbf{q})$, can be written as (21)

$$F(\mathbf{q}) = \sum\nolimits_i f_i(q) B_{fi} \exp(-\mathbf{q}\cdot\mathbf{r}_i),$$

where $\mathbf{r}_i$ is the vector from the origin of the unit cell to the average position of the ith atom, and

$$B_{fi} = \exp\left(-B_i q^2/16\pi^2\right).$$

If the distributions of the displacements of atoms from their average positions, $\boldsymbol{\delta}_i$, are isotropic, then

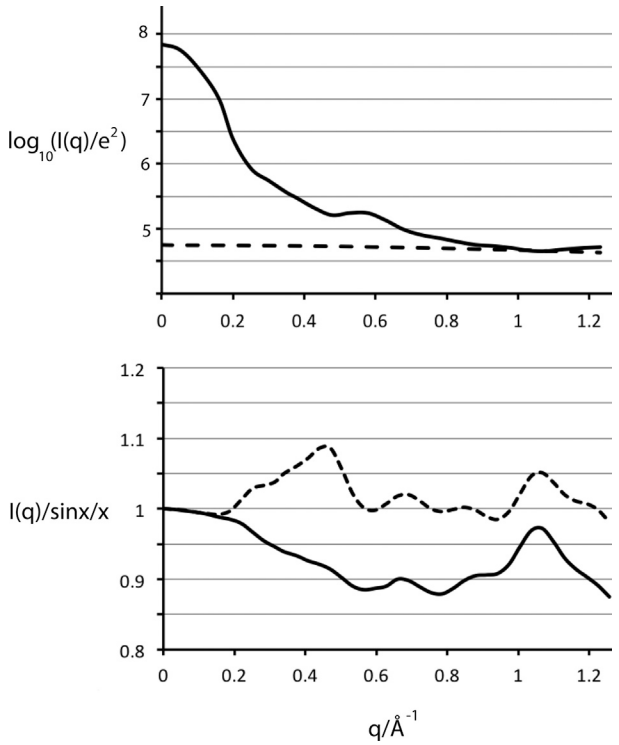

FIGURE 2 The effect of correlated motion on solution-scattering profile of the A-chain of Φ6 lysin. (*Top panel*) The solution-scattering profile for the A-chain of Φ6 lysin (PDB:4DQ5) (26) computed using the Debye equation (Eq. 1), which takes no account of atomic displacements. No hydrogen atoms were included in the computation, and for each region of the chain where alternative conformations are proposed in PDB:4DQ5, only the *A* conformation was taken into account. The data (*solid line*) are plotted as $\log_{10}(I)$ vs. q. Also shown is the profile that corresponds to $\sum_i f_i(q)^2$, which is plotted the same way (*broken line*). (*Bottom panel*) This figure illustrates the impact that different models for the covariations of atomic displacements can have on the solution-scattering profiles predicted for macromolecules. A profile was computed for Φ6 lysin using Eq. 5, and assigning to each atom an isotropic B-factor twice that reported in PDB:4DQ5 to ensure that displacement effects would be obvious. The average B-factor for all heavy atoms in the molecule was ~45 Å^2, which is by no means unusual for structures deposited in the PDB. (*Solid lines*) Profile from using Eq. 5 divided by the profile using Eq. 1 (see *top panel*). Using the TLS model proposed for Φ6 lysin in PDB:4DQ5, but again doubling all B-factors, a profile was computed for Φ6 lysin using Eq. 4. (*Dashed-lines*) Profile from using Eq. 4 divided by the profile using Eq. 1.

$$B_i = 8\pi^2\langle\delta_i^2\rangle/3$$

and

$$B_{fi} = \exp\bigl(-q^2\langle\delta_i^2\rangle/6\bigr).$$

(These expressions are easily modified to allow for anisotropic displacements.) The B_i are called temperature factors, or B-factors. Estimates of B-factors are obtained for crystal structures as they are refined, and that information is included in the coordinate files. The term

$$\exp\left(-\frac{1}{2}q^2\mathbf{C}_{ij}^T\bigl(\mathbf{V}_i+\mathbf{V}_j-2\mathbf{V}_{ij}\bigr)\mathbf{C}_{ij}\right)$$

in Eq. 4 is the solution-scattering equivalent of a crystallographic B-factor correction.

Although the B-factors reported for crystal structures can be useful, they do not contain all the information required for the prediction of solution-scattering profiles,

$$\boldsymbol{\Delta}_{ij} = \boldsymbol{\delta}_i - \boldsymbol{\delta}_j.$$

It follows that

$$\langle\boldsymbol{\Delta}_{ij}^2\rangle = \langle\boldsymbol{\delta}_i^2\rangle - 2\langle\boldsymbol{\delta}_i\cdot\boldsymbol{\delta}_j\rangle + \langle\boldsymbol{\delta}_j^2\rangle$$

where $\langle\boldsymbol{\delta}_i{}^2\rangle$ and $\langle\boldsymbol{\delta}_j{}^2\rangle$ are the variances of the displacements of the two atoms, and $\langle\boldsymbol{\delta}_i\cdot\boldsymbol{\delta}_j\rangle$ is the corresponding covariance. The $\langle\boldsymbol{\delta}_i\cdot\boldsymbol{\delta}_j\rangle$ values have no effect whatsoever on the intensities of Bragg reflections, and hence cannot be extracted from crystallographic data. Nevertheless, diffraction patterns do contain information about covariances. Covariation between the thermal motions of atoms in crystals contributes to the diffuse scatter observed in diffraction patterns between Bragg reflections (21). Unfortunately, as of this writing, no one yet has discovered a method for recovering covariances from diffuse scattering patterns (see Moore (22)).

It is important to recognize that the B-factors crystallographers report, and variances relevant to solution scattering, are not directly comparable. The atomic coordinates reported by crystallographers are measured with respect to the axes of the pertinent unit cell, rather than some axis system anchored in the molecule to which those atoms belong. Thus not only will lattice disorder contribute to crystallographic B-factors, but so too will whole-molecule translations and rotations, which are irrelevant for solution-scattering experiments. In addition, crystal-packing interactions may affect the range of motions that are possible for a molecule. Nevertheless, for lack of anything better, one might consider using the highly simplified version of Eq. 4 that follows and crystallographic B-factors to obtain predictions for solution-scattering profiles that take account of conformational fluctuations to a first-order of approximation:

$$I(q) = \sum_i\sum_j f_i(q)f_j(q)\exp\bigl(-\bigl(B_i+B_j\bigr)q^2/16\pi^2\bigr) \times \sin\bigl(qr_{ij}\bigr)/\bigl(qr_{ij}\bigr). \quad (6)$$

(Note that the temperature factors appropriate for all the self terms, (i,i), in Eq. 5 are zero; note also that this equation is an approximate representation of the solution-scattering profile that would be obtained for a macromolecule, the electron density distribution of which is the same as that of the average of all the molecules in the crystals used to determine its structure.) Expressions similar to Eq. 5 have been used in the past (23,24), and in at least one instance, derived using an approach similar to the one presented here (19). If the fluctuations in atomic positions are isotropic

and do not co-vary, and the B-factors reported for some macromolecular crystal structure are relevant in solution, Eq. 5 will be exact to fourth-order.

It is important to recognize that predictions made using Eq. 5 are likely to be inaccurate. The atoms in the interiors of proteins are packed almost as densely as the atoms in an organic solid (25). Thus, not only will the thermal motions of an atom in the interior of a protein correlate with those of the atoms to which it is covalently bonded, they will also necessarily correlate with those of the atoms' nonbonded neighbors.

Correlated motions make a difference

It has already been shown that thermal displacements can significantly affect the contributions made to solution-scattering profiles by pairs of atoms. What effect do they have on the scattering profiles of entire macromolecules? Could it be, for example, that the Debye equation works reasonably well, despite its shortcomings, because errors made in computing the contribution of one pair of atoms are cancelled out by the errors made for other pairs?

This issue was addressed computationally using the structure of a small viral protein called Φ6 lysin (monomer molecular mass ~ 17 kDa) as the test object (PDB:4DQ5) (26). In its crystals, this protein forms an asymmetric homodimer. The computations described below were done using the structure reported for only one of the two chains of that dimer, and it was assumed that the crystallographic B-factors reported for that chain are applicable in solution.

One reason this protein was used as a test molecule is that its crystal structure has been refined using translation/libration/screw (TLS) methods (27–30). In a TLS refinement, macromolecules are treated as assemblies of domains, each of which rotates and translates thermally, independently of its neighbors. These rigid body motions not only make the thermal motions of the atoms in each domain anisotropic, they also make them highly correlated. The B-factor assigned to an atom in a macromolecular crystal structure that has been refined this way is the sum of the anisotropic B-factor imparted to it by the rigid body motions of its domain, and an isotropic B-factor that represents its thermal motion relative to its domain. Thus, for this particular protein, not only are there anisotropic B-factor estimates available, there is also a model for the correlations between the thermal motions of its atoms. It follows that the solution-scattering profile of this particular protein could be computed using all of the equations of interest here: Eq. 1 (no thermal motions), Eq. 5 (uncorrelated thermal motions that are isotropic), and Eq. 4 (correlated thermal motions that are anisotropic). In addition, a profile was computed for the protein that used the anisotropic B-factors available, but ignored their correlations.

The top panel of Fig. 2 shows the solution-scattering profile that the Φ6 lysin monomer would give if its atoms were all fixed at their average locations, plotted as $log_{10}(I(q))$ versus q (*solid line*). It is the protein's Debye equation, i.e., Eq. 1, profile. This panel also shows the component

$$\sum_i f_i(q)^2$$

of that profile plotted the same way (*broken line*). The two profiles converge at high scattering angles, as expected (see Wilson (31)).

The bottom panel (*solid line*) is the Eq. 5 profile of the Φ6 lysin monomer divided by its Eq. 1 profile (Fig. 2, *top panel*). The average B-factor assigned to the atoms in the structure was ~45 $Å^2$, which corresponds to an average value of ~0.57 $Å^2$ for the variance of the displacements of (heavy) atoms in any given direction. Equation 5 takes no account of correlations between the motions of different atoms, and because it is impossible that the motions of atoms in a molecule could ever be entirely uncorrelated, the profile it predicts is best thought of as the profile that would be approached in the limit as correlations go to zero. Interestingly, while the Eq. 5 profile lies below the Eq. 1 profile for all $q > 0$, and decreases more or less steadily out to $q \sim 0.6$, because at higher scattering angles it fluctuates. The profile obtained by assuming that the B-factors for the test protein are anisotropic, but uncorrelated, was nearly identical to that obtained using Eq. 5, which assumes that all B-factors are isotropic (data not shown).

The solution-scattering profile of a macromolecule subject to TLS disorder is readily computed. For the atom pairs within each domain, only the extra isotropic components of their B-factors need to be taken into account because interatomic distances within a domain are unaffected by rigid body translations and rotations, and by definition, the extra components of those B-factors are uncorrelated. For pairs of atoms that belong to different domains, entire anisotropic B-factors must be used, but for those pairs the $\mathbf{V}_{ij}$ are all zero because, by assumption, the rigid body motions of the different domains of the macromolecule do not correlate. The dotted-line profile in the bottom panel of Fig. 2 is the TLS profile of the Φ6 lysin monomer divided by its Eq. 1 profile.

Because the average B-factor assigned to atoms in the TLS computation is less than that for the no-correlation computation, it was to be expected that the TLS curve would be everywhere more positive than the no-correlation curve. The surprise is that for all $q > \sim 0.20$ $Å^{-1}$, the intensity predicted using the TLS model for the thermal disorder of its atoms is actually greater than it would be if there were no thermal disorder whatever.

Computational experiments done using these data demonstrated that the differences between the two profiles shown in the bottom panel of Fig. 2 are entirely to due to differences in the way variance/covariance information available is treated, and the impact such treatment will subsequently have on the $\sin x/x$ part of the first double-sum term on the right-hand side of Eq. 4. (The absolute magnitude of the contribution made by the A part of that double

sum is everywhere <1% of the total predicted intensity, and the contribution made by the second double sum is so small it could have been omitted entirely.) The reason the TLS profile is larger than the Eq. 1 profile appears to be that the vast majority of the short interatomic distances in the Φ6 lysin monomer are intradomain distances, and hence the B-factor corrections applied to those distances tend to be significantly smaller in the Eq. 4 calculation than they are in the Eq. 5 calculation. Because most of the long distances in the protein involve atoms belonging to different domains, the B-factor corrections applied to them are the same in both calculations. The systematic downweighting of long-distance contributions relative to short distance contributions in the Eq. 7 calculation systematically distorts the predicted scattering profile in the manner shown.

The scattering angle at which thermal disorder begins having appreciable effects is inversely related to average B-factor

The scattering angle at which thermal disorder begins to have a measurable impact on solution-scattering profiles is determined by the (average) magnitude of that disorder, not the structure of the molecule displaying it. If effects that alter intensities by 1% are taken as being detectable, then Eq. 4 profiles should begin deviating significantly from the Eq. 1 profiles when

$$0.99 = \exp\left(-V_{\text{avg}}q^2/2\right),$$

which will be true when

$$q = \sqrt{0.02/V_{\text{avg}}},$$

where $V_{\text{avg}} = B_{\text{avg}}/8\pi^2$ and B_{avg} give the average, isotropic B-factor of the heavy atoms in the macromolecule. For the Φ6 lysin protein structure discussed above, B_{avg} was set to 45 Å^2, and thus, for this molecule, deviations should have become apparent starting at $q \sim 0.19$ Å^{-1}, which is approximately what is seen in Fig. 2.

This analysis implies that thermal disorder is unlikely to have an appreciable impact on radius-of-gyration estimates. The radius of gyration of a macromolecule is determined by the second-order term in Eq. 2 (32). If there is no thermal disorder, its value will be

$$-(1/3!)\sum_i\sum_j f_i(q)f_j(q)q^2R_{ij}^2,$$

but if there is thermal disorder, it will be

$$-(1/3!)\sum_i\sum_j f_i(q)f_j(q)q^2\left(R_{ij}^2 + \left\langle\Delta_{ij}^2\right\rangle\right).$$

It follows that thermal disorder makes the radius of gyration estimates obtained from ensembles of macromolecules larger than that of their average members, but the effect is very small. If the average value of R_{ij} is ~20 Å, which it would be for a protein of modest molecular weight, and the average value of $\langle\Delta^2{}_{ij}\rangle$ is ~1.0 Å, which would also be unexceptional, then the radius-of-gyration estimate obtained by inserting crystallographic coordinates into Eq. 1 would differ from the measured radius of gyration by only a few hundreds of an Ångstrom—an amount so small it would be impossible to measure reliably.

DISCUSSION

The equations developed above may appear quite general, but in fact, they have significant limitations. To begin with, they can only be applied when the structure of the average member of a solution of chemically identical macromolecules is known at atomic resolution. This requirement is unlikely to be met if the molecule of concern adopts two or more distinctly different conformations in solution—simply because of the prevailing lack of information about relative populations if nothing else. Furthermore, if a macromolecule behaves this way in solution, it will be difficult to obtain meaningful estimates for the variances and covariances of its atomic displacements, some of which, at least, are likely to be very large. This is important because, as pointed out earlier, Eq. 4 is a truncated version of an infinite series. It can only be expected to give accurate results if the solution-scattering profile sought is not far from that of the average molecule, and that is unlikely to be the case if the conformational heterogeneity that must be dealt with is very large.

The use of functions of the form $\exp(-Dq^2)$ in Eq. 4 also imposes limitations. These functions are approximations to power series of the form

$$1 + A\langle\text{arg}\rangle + B\left\langle\text{arg}^2\right\rangle + \ldots.$$

If the distribution functions over which atomic displacements are averaged are Gaussians, they will be accurate for all values of q. If the distributions functions are not Gaussian, but are nevertheless even functions, then $\exp(-Dq^2)$ will represent the corresponding power series well only through its cubic term. When even this condition is unmet, Eq. 4 will be usefully accurate only for very small values of Dq^2, say for $Dq^2 < 0.25$. This condition will again make Eq. 4 useful only for macromolecules that are fluctuating thermally within a single well in their conformational free energy landscapes.

It will be impossible to make much practical use of the expressions developed above unless reliable information can be obtained about the variances and covariances of the thermal motions of the atoms in macromolecules. The most crystallography can at provide is an indication about the magnitudes of the elements of the variance/covariance matrices of individual atoms (see above). However, only for crystals solved to quite high resolutions, where the ratio of observations made to parameters determined is very high,

are the B-factors in the PDB likely to describe the disorder in the relevant crystals accurately, and even then there will be no information available about atom-atom covariances. The problem with B-factors is the procedures used to refine crystal structures that have been solved at lower resolutions tend to make B-factor estimates the repository for errors of all kinds. It appears that for the moment, molecular dynamics simulations may be the best source of information we have about variances and covariances. Although they are not without their own shortcomings, there is good reason to believe that they will improve in the future.

In conclusion, thermal disorder poses challenges to those interested in extracting information from comparisons between experimental S/WAXS profiles and predicted solution-scattering profiles. For example, how is one to distinguish among differences between predicted and measured profiles caused by shortcomings in the treatment of solvent from those caused by inadequacies in the way thermal disorder has been modeled, or even from a lack of correspondence between crystal structures and average structures in solution? It is hoped that the analysis provided above will help motivate the single- and wide-angle x-ray scattering communities to address these problems.

SUPPORTING MATERIAL

A series of supporting equations, with explanatory text, is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00221-5.

I thank Professor John Tainer and Dr. Robert Rambo for the insightful comments they provided on an early draft of this paper.

REFERENCES

1. Glatter, O., and O. Kratky, editors. Small Angle X-Ray Scattering Academic Press, London.
2. Graewert, M. A., and D. I. Svergun. 2013. Impact and progress in small and wide angle x-ray scattering (SAXS and WAXS). *Curr. Opin. Struct. Biol.* 23:748–754.
3. Mertens, H. D. T., and D. I. Svergun. 2010. Structural characterization of proteins and complexes using small-angle x-ray solution scattering. *J. Struct. Biol.* 172:128–141.
4. Rambo, R. P., and J. A. Tainer. 2010. Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle x-ray scattering. *Curr. Opin. Struct. Biol.* 20:128–137.
5. Rambo, R. P., and J. A. Tainer. 2013. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature.* 496:477–481.
6. Debye, P. 1915. X-ray dispersion [Zerstreuung von Röntgenstrahlen]. *Ann. Phys.* 46:809–823.
7. Svergun, D., C. Barberato, and M. H. J. Koch. 1995. CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* 28:768–773.
8. Park, S., J. P. Bardhan, …, L. Makowski. 2009. Simulated x-ray scattering of protein solutions using explicit-solvent models. *J. Chem. Phys.* 130:134114–134118.
9. Schneidman-Duhovny, D., M. Hammel, and A. Sali. 2010. FOXS: a web server for rapid computation and fitting of SAXS profiles. *Nucl. Acids Res.* 38:W540–W544.
10. Grishaev, A., L. Guo, …, A. Bax. 2010. Improved fitting of solution x-ray scattering data to macromolecular structures and structural ensembles by explicit water modeling. *J. Am. Chem. Soc.* 132:15484–15486.
11. Poitevin, F., H. Orland, …, M. Delarue. 2011. AQUASAXS: a web server for computation and fitting of SAXS profiles with non-uniformly hydrated atomic models. *Nucl. Acids Res.* 39:W184–W189.
12. Liu, H., R. J. Morris, …, P. H. Zwart. 2012. Computation of small-angle scattering profiles with three-dimensional Zernike polynomials. *Acta Crystallogr. A.* 68:278–285.
13. Makowski, L., D. Gore, …, R. F. Fischetti. 2011. X-ray solution scattering studies of the structural diversity intrinsic to protein ensembles. *Biopolymers.* 95:531–542.
14. Bernadó, P., E. Mylonas, …, D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.
15. Pelikan, M., G. L. Hura, and M. Hammel. 2009. Structure and flexibility within proteins as identified through small angle x-ray scattering. *Gen. Physiol. Biophys.* 28:174–189.
16. Bernadó, P. 2010. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur. Biophys. J.* 39:769–780.
17. Yang, S., L. Blachowicz, …, B. Roux. 2010. Multidomain assembled states of HCK tyrosine kinase in solution. *Proc. Natl. Acad. Sci. USA.* 107:15757–15762.
18. Makowski, L., J. P. Bardhan, …, R. F. Fischetti. 2011. WAXS studies of the structural diversity of hemoglobin in solution. *J. Mol. Biol.* 408:909–921.
19. James, R. W. 1932. On the influence of temperature on the scattering of x-rays by gas molecules [Über den einfluss der temperatur auf die streuung der Röntgenstrahlen durch gasmoleküle]. *Physikal. Zeit.* 33:737–754.
20. Busing, W. R., and H. A. Levy. 1964. The effect of thermal motion on the estimation of bond lengths from diffraction measurements. *Acta Crystallogr.* 17:142–146.
21. James, R. W. 1965. The Optical Principles of the Diffraction of X-Rays. Cornell University Press, Ithaca, NY.
22. Moore, P. B. 2009. On the relationship between diffraction patterns and motions in macromolecular crystals. *Structure.* 17:1307–1315.
23. Zhang, R., P. Thiyagarajan, and D. M. Tiede. 2000. Probing protein fine structure by wide angle solution x-ray scattering. *J. Appl. Cryst.* 33:565–568.
24. Tiede, D. M., R. Zhang, …, J. S. Lindsey. 2004. Structural characterization of modular supramolecular architectures in solution. *J. Am. Chem. Soc.* 126:14054–14062.
25. Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.
26. Dessau, M., D. Goldhill, …, Y. Modis. 2012. Selective pressure causes an RNA virus to trade reproductive fitness for increased structural and thermal stability of a viral enzyme. *PLOS Gen.* 10.1371/journal.pgen.1003102.
27. Cruickshank, D. W. J. 1956. The analysis of the anisotropic thermal motion of molecules in crystals. *Acta Crystallogr.* 9:754–756.
28. Schomaker, V., and K. N. Trueblood. 1968. On the rigid-body motion of molecules in crystals. *Acta Crystallogr. B.* 24:63–76.
29. Kuriyan, J., and W. I. Weis. 1991. Rigid protein motion as a model for crystallographic temperature factors. *Proc. Natl. Acad. Sci. USA.* 88:2773–2777.
30. Winn, M. D., M. N. Isupov, and G. N. Murshudov. 2001. Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.* 57:122–133.
31. Wilson, A. J. C. 1949. The probability distribution of x-ray intensities. *Acta Crystallogr.* 2:318–321.
32. Guinier, A., and G. Fournet. 1955. Small-Angle Scattering of X-Rays. John Wiley & Sons, New York.